基于决策树模型的区域$PM_{2.5}$污染管控时空识别
——以关中地区为例

贾 册[1]，陈 臻[2,3]，韩 梅[4]

（1. 中国人民大学环境学院，北京 100872；2. 中国科学院科技战略咨询研究院，北京 100190；
3. 中国科学院大学公共政策与管理学院，北京 100049；4. 陕西省环境调查评估中心，陕西 西安 710054）

摘 要： 以关中地区为研究区域，基于时空聚类和决策树模型提出一种简易的$PM_{2.5}$污染管控时空识别方法。首先使用时空聚类算法对冬防期$PM_{2.5}$浓度进行聚类，识别不同的$PM_{2.5}$污染区域，基于不同区域的气象数据分别构建决策树模型，识别不同区域影响$PM_{2.5}$浓度最不利扩散的气象条件，分析最不利气象条件下的$PM_{2.5}$浓度变化情况，以此确定各区域需要进行污染管控的时间段。结果表明：(1) 时空聚类方法识别出关中地区$PM_{2.5}$分布主要呈现出低海拔平原区域和海拔相对较高的山脉区域。(2) 决策树模型分析结果显示：高海拔区域在Ⅰ-10（1.57 h≤日照时数<7.88 h、最大风速<3.72 $m\cdot s^{-1}$）和Ⅰ-11（日照时数<1.57 h、最大风速<3.72 $m\cdot s^{-1}$）两类气象条件下，区域的$PM_{2.5}$浓度保持较高水平；低海拔区域在Ⅱ-10（小型蒸发量≥0.96 mm、平均相对湿度≥45.38%、日照时数<8.55 h、平均风速≥2.43 $m\cdot s^{-1}$）和Ⅱ-11（小型蒸发量<0.96 mm）两类气象条件下，区域的$PM_{2.5}$浓度保持较高水平。(3) 回归结果显示，关中地区低海拔区域和高海拔区域在最不利气象条件下，$PM_{2.5}$浓度平均会持续上升4.76 d，直至最高浓度。

关键词： 决策树模型；$PM_{2.5}$；分区管控；重污染

《2020中国生态环境状况公报》显示，2020年全国337个城市累计发生大气环境严重污染345 d，重度污染1152 d，可见当前中国大气环境污染形势仍然很严峻。这些重污染天气的出现不仅影响居民身心健康，还对社会经济发展造成了很大的损失[1-5]。重污染天气出现的核心原因是由于气象条件的变化，使得大气环境中的污染物不能被及时清除而累积形成。相关研究表明，在重污染天气期间进行预警能有效降低人的出行，减少移动源等污染物的排放从而达到减缓重污染的程度[6-10]。当前国内应对重污染天气的过程中也采取了预警的方式来应对。根据空气质量情况，地方政府适时启动红色、橙色、黄色级别的重污染天气应急预警，要求工业源、移动源等污染排放在一定时间内做出不同的污染物减排，实现空气质量的改善[11-12]。国内外进行污染预警的过程均是通过模型进行预测未来24 h、48 h甚至一周的空气质量变化情况来随时调整预警的级别。这种不确定预警时间和预警时长的方式严重干扰了企业的正常生产，据统计咸阳市2019年1—3月和11—12月期间共预警76 d，占到总天数的一半。当前污染预警管控的过程中是以地级市政府为主体发布预警，要求辖区内企业进行相应的污染物减排，而部分地级市存在行政区过大，同一时间内的$PM_{2.5}$可能存在两极分化的情况。

由于重污染天气发生一般是区域性的[13]，重污染区域内应该统一进行管控，并有效识别需要进行管控的时间段。对于确定的区域而言，短时间内对其大气污染物环境容量影响最大的是气象因素的变化[14-17]，高爽等[18]提出基于气象分型的大气环境容量测算，但该研究是通过人工对气象进行分类识别，该方式存在划分依据受人为主观影响的问题，现有研究气象和排放对环境中$PM_{2.5}$浓度变化的相关文献，基本没有突破行政区划来体现污染区域，也没有深入挖掘所分类气象条件下的污染物变化

收稿日期： 2022-01-18； **修订日期：** 2022-04-23
基金项目： 中国人民大学科学研究基金（中央高校基本科研业务费专项资金资助）项目成果（No.21XNH057）；陕西省重点研发计划（2021SF-501）
作者简介： 贾册（1994-），男，博士研究生，主要从事环境政策与管理. E-mail: adam-jia@qq.com
通讯作者： 韩梅. E-mail: 158697694@qq.com

规律,无法有效指导在不利气象条件下污染排放管控[19-20]。故需要相关研究来对$PM_{2.5}$的管控时空进行识别,从而实现精确管控和预警,使工业企业提前做好应对,减少经济损失。基于此,本文基于决策树模型提出了一种简易的区域$PM_{2.5}$污染管控时间段识别方法,以期有效指导重污染天气期间$PM_{2.5}$污染管控。

1　研究方法

1.1　研究思路

$PM_{2.5}$污染现象主要出现在冬防期(11月至次年3月)[21],本文基于冬防期$PM_{2.5}$的浓度数据并结合地形等数据对所研究区域在时间纬度上进行空间聚类,识别不同的$PM_{2.5}$污染区域;基于$PM_{2.5}$浓度数据对不同区域的气象数据分别构建决策树模型,识别不同区域影响$PM_{2.5}$浓度最不利扩散的气象条件,对最不利扩散气象条件下的$PM_{2.5}$变化情况进行模型构建,确定不同研究区域的$PM_{2.5}$管控时间段(图1)。

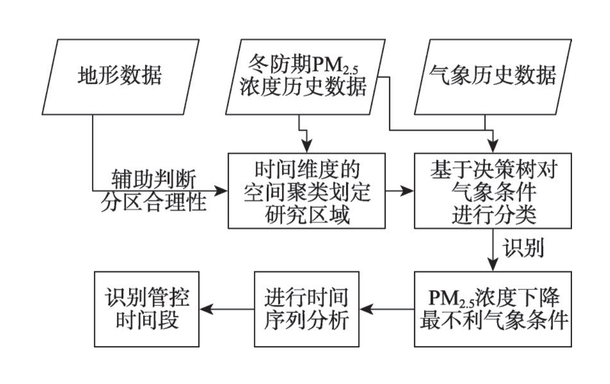


图1　研究思路

Fig. 1　Research idea

1.2　基于时间维度的污染区域划定

受地形、经济发展、产业布局和气象条件等因素的影响,污染区域分布往往与行政区划不一致,进行区域污染联防联控尤为重要。本文所研究的关中地区(西安市、咸阳市、渭南市、铜川市、宝鸡市、韩城市及杨凌区)具有相对独立的地理环境,包括关中平原和围绕着关中平原的北山山系、秦岭以及崤山等山脉,具有相对独立的地形气候[22-23]。基于此,本文在进行污染区域的识别和划定过程中,综合考虑时空纬度,识别不同气象条件下需要协同管控的污染区域。

首先是对所研究区域每日空气质量监测站点的$PM_{2.5}$浓度进行空间聚类,将每日聚类的分区结果进行相关性聚类分析,识别在研究时间段内应该集中进行管控的区域。对于所识别的区域,在排除外来传输污染的情况下,工业源、移动源、无组织面源等的污染减排,都将会降低该区域的$PM_{2.5}$浓度。在进行聚类过程中先由监测点生成泰森多边形,对生成的泰森多边形进行聚类分析,以保证所识别污染区域的连续性。

1.3　决策树模型

分类算法中比较常见的有决策树(分类树)、随机森林、支持向量机、神经网络分类等算法,其中决策树算法能够直观地给出详细的分类过程,其他模型更偏向于黑箱模型预测。决策树算法由Leo Breiman等学者在1984年提出[24],并发展出C4.5、C5.0等算法模型。决策树的表现形式为二叉树,在进行模型构建的过程中可分为生长和剪枝2个过程。生长的过程中主要是对输入的众多变量中选取最佳变量以及对所选取的最佳变量寻找最佳分割点。剪枝的过程是找到最佳变量和分割点后将其他影响模型精度的树枝减掉。其中,最佳的分组变量和分割点是被分类变量异质性下降最快的对应参数。该模型能够从繁杂的数据指标中有效识别最有利于划分$PM_{2.5}$浓度的气象指标因子及相关指标阈值。

1.4　数据来源

研究时段为2016—2019年间的冬防期,每年11月1日至次年3月31日。所使用的$PM_{2.5}$监测数据来自关中五市的国控点及省控点环境空气质量监测站,共87个有效监测点,空气质量数据来源于陕西省空气质量实时发布系统。气象数据来源于中国气象数据网(http://data.cma.cn),中国地面气候资料日值数据集(V 3.0)包括平均风速、最大风速、平均本站气压、日最高本站气压、日最低本站气压、平均气温、日最高气温、日最低气温、平均地表气温、日最高地表气温、日最低地表气温、平均相对湿度、是否降水、累计降水量、小型蒸发量、日照时数等指标。

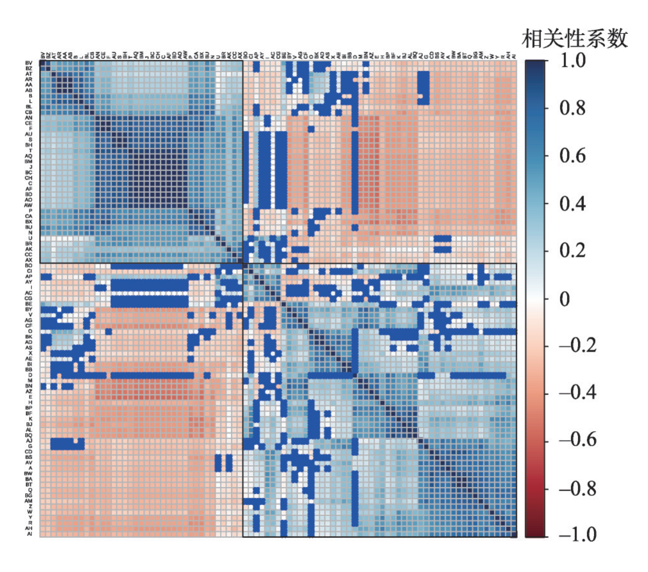
2　结果与分析

2.1　聚类分析

关中区域在研究时间段有87个空气质量监测点,基本上保证了关中各区县有1个空气质量监测

点，通过对所有监测点的2016年11月至2019年3月的1—3月、11—12月期间每日的$PM_{2.5}$浓度值进行空间聚类，共生成453个聚类结果。聚类过程中采用R语言的rgeoda包K均值聚类方法进行空间聚类。在对关中地区进行聚类的过程中，通过选择不同的聚类系数并结合F-Stat、地形等因素后，发现关中地区在进行$PM_{2.5}$浓度空间聚类的过程被分成两类最优的聚类。对453个聚类的结果求相关性系数识别需要集中管控的区域，相关性系数结果见图2。将相关性系数进行聚类可将87个空气质量监测点划分为两类（图2中黑色方框所包含的区域），这些点在整个研究时间段内会呈现高度一致性，可在地理上划分在一个区域中进行集中管控。在条件允许下，可进一步细化分区管控区域，本文将关中地区分为两类进行研究。将图2分区后的监测点生成泰森多边形，整体区域边界采用函数将边缘进行光滑处理在行政区地图中进行体现（图3a），区域一（白色区域）对应空气质量监测点有37个点，气象观测点有7个点，$PM_{2.5}$浓度平均值为116.55 μg·m^{-3}，超过24 h $PM_{2.5}$标准值的天数为192 d，占总天数的42.38%。区域二（灰色区域）对应空气质量监测点有50个点，对应的气象观测点有4个点；$PM_{2.5}$浓度平均值为112.05 μg·m^{-3}，超过24 h $PM_{2.5}$标准值的天数为294 d，占总天数的64.90%。

将分区结果投影至关中地区的高程图，由图3b可以看出，区域二划定的范围基本与关中平原所在位置重叠，可见关中地区处于低海拔区域的地方有非常强的区域性[25]，同时也说明了$PM_{2.5}$浓度变化与地形有较大的关系，这与张波等[26]的研究基本一致，有效证明了当前分区的合理性。



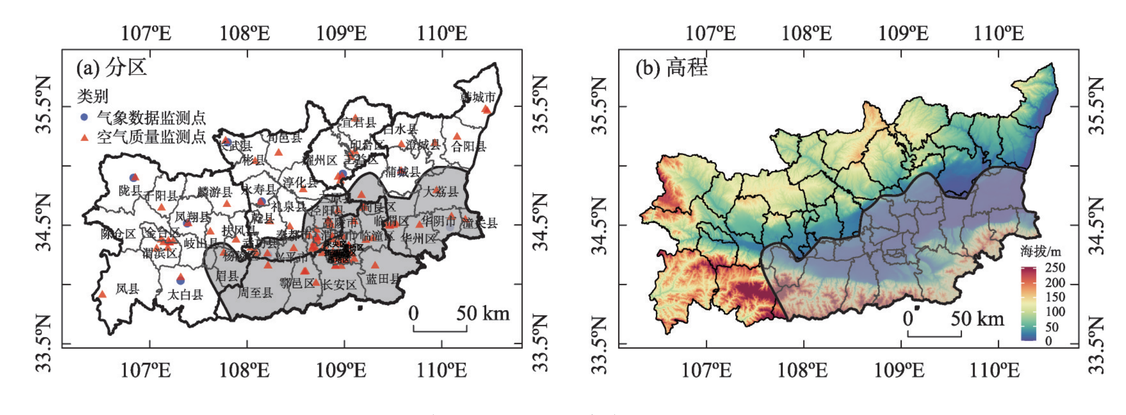
注：图中每个方框颜色深浅代表的是该监测点和其他监测点在研究时段内每日的聚类结果相关性系数；蓝色“×”号代表二者的相关性未通过99%显著性检验。

图2　关中地区空气质量监测点$PM_{2.5}$浓度聚类结果相关性

Fig. 2　Correlation of concentration clustering results at $PM_{2.5}$ in Guanzhong area

2.2　地形影响分析

综上所述，关中地区的$PM_{2.5}$浓度变化与地形有很大的关系，将研究时段内各环境空气质量监测点的$PM_{2.5}$浓度平均值采用反距离权重（IDW）方法进行空间差值，按照投影到东西、南北和高程3个维度，进行趋势分析（图4）。结果表明，从东到西$PM_{2.5}$浓度呈先增加后下降并逐步趋于平稳，在经度109°左右达到了最高点（图4a）。从南到北而言，$PM_{2.5}$浓度呈先上升后下降再上升的变换趋势，在纬度34.3°左右达到了最高点（图4b）。从海拔来看，$PM_{2.5}$浓度



注：白色区域为区域一，灰色区域为区域二。

图3　关中地区$PM_{2.5}$浓度地区聚类分布

Fig. 3　Regional cluster distribution of $PM_{2.5}$ concentration in Guanzhong area

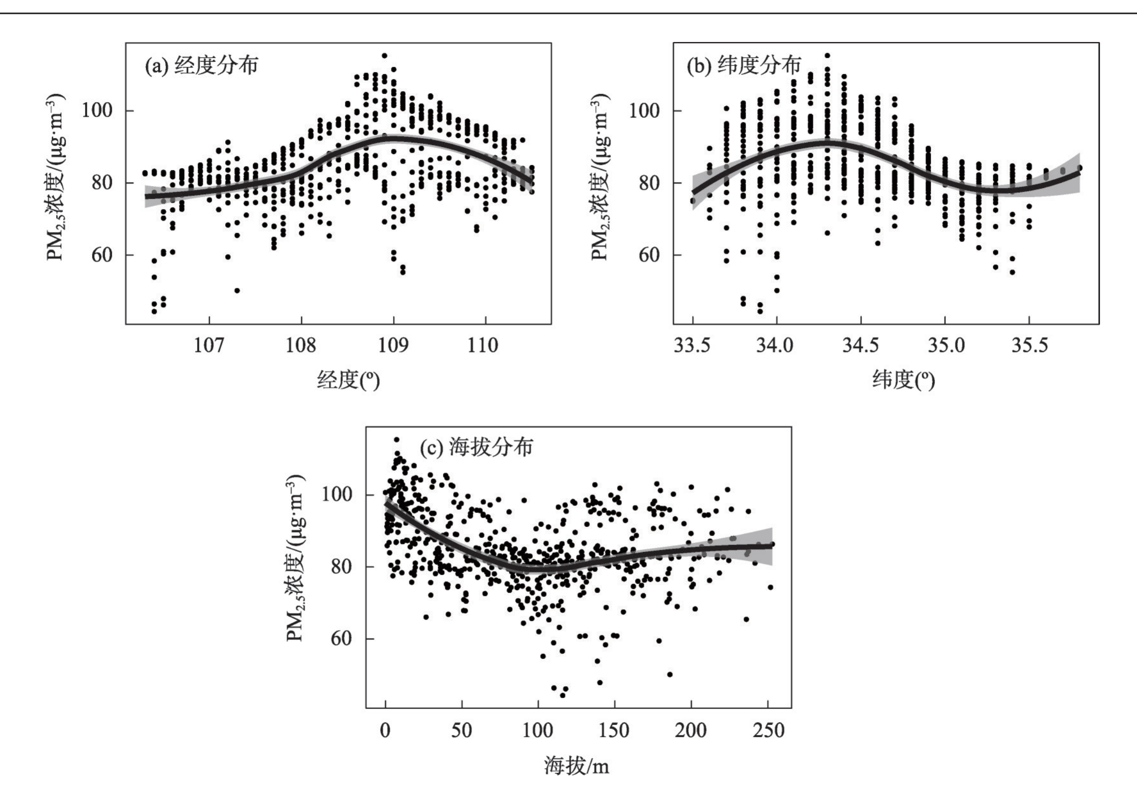


图4　$PM_{2.5}$浓度分布与地形关系分布

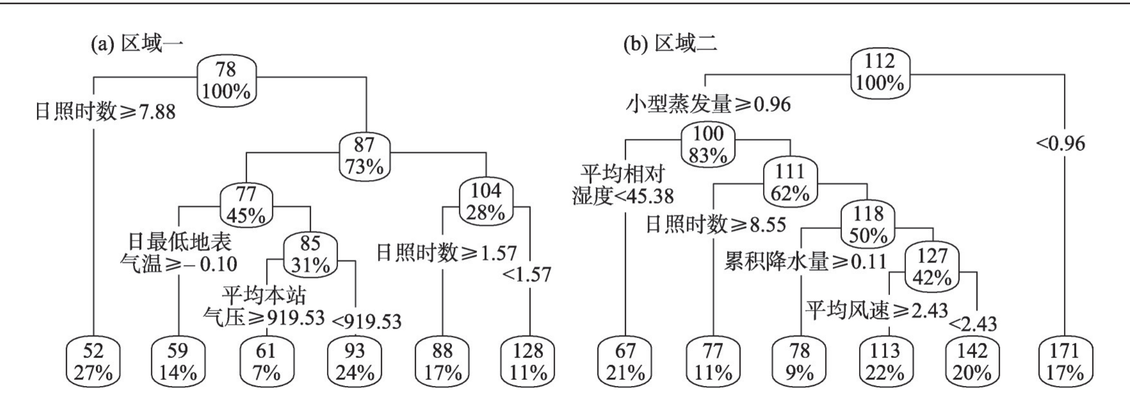
Fig. 4　Relationship between $PM_{2.5}$ concentration distribution and topography

从低海拔地区随着海拔上升的过程中先下降,当海拔达到100 m左右之后,$PM_{2.5}$浓度基本趋于稳定,在85 μg·m^{-3}(图4c)。结合图3a可以看出,关中地区$PM_{2.5}$浓度最高的区域集中在西安市城区,$PM_{2.5}$浓度从西安市城区向四周的西咸新区、咸阳市、渭南市、铜川市的平原(低海拔)地区逐步过渡,形成较强区域性的污染区域。随着四周海拔的不断上升,$PM_{2.5}$浓度又开始出现下降的情况,当海拔达到一定高度后,海拔将不在是影响$PM_{2.5}$浓度变化的因素,更多是地区污染排放在影响其变化。

2.3　决策树模型下气象分类结果

由于地形和污染企业数量在一定时间范围内是不会发生特别大的变化,而气象条件随时在变化,主要研究气象对$PM_{2.5}$浓度变化的影响。由于$PM_{2.5}$浓度的变化是污染排放和多项气象因子综合作用的结果,人工对气象因子进行判断和划定相关阈值会导致较大的主观不确定性。为了规避人为主观因素的影响,采用较为成熟的决策树模型来识别不同区域影响$PM_{2.5}$浓度的关键气象因子[27]。决策树模型可以通过气象因子对$PM_{2.5}$浓度变化情况进行分割,在分割过程中充分考虑$PM_{2.5}$浓度和气象因子之间的关系,并将具有共线性的气象因子进行剔除,有效减少干扰因子。同时该方法能够通过图形表达,更直观地对分类过程给出依据,相较于其他黑箱决策树模型有更好地解释性。

故采用决策树模型来识别影响$PM_{2.5}$浓度的关键气象因子和分类阈值。在识别过程中以$PM_{2.5}$浓度为y,所有气象因子为x,采用rpart函数进行分类,初始分类结果结合复杂度参数最佳cp值进行剪枝,结果见图5。其中,区域一在模型构建的时候选取日照时数、最大风速、日最低地表气温、平均本站气压4项指标对气象进行分类,针对$PM_{2.5}$浓度的变化情况将气象分为6类。区域二在模型构建的时候选取小型蒸发量、平均相对湿度、日照时数、累积降水量、平均风速5项指标进行分类,针对$PM_{2.5}$浓度的变化情况将气象分为6类。区域一、二的气象分类结果见表1,区域一、二中均识别出日照时数的变化主要是由于$PM_{2.5}$污染造成的。风速对于$PM_{2.5}$的去除有非常大的效果,也被模型所识别,其中日最大风速和日平均风速是高度相关可替代的因子。除日照时数和风速因子外,区域一处于海拔相对较高地区,还识别了日最低地表气温、平均本站气压作为关键变量,而区域二处于低海拔平原地区,并且$PM_{2.5}$浓度严重程度远高于区域一,在不考虑排放的



注：图中单位省略。

图5　$PM_{2.5}$浓度气象决策树

Fig. 5　$PM_{2.5}$ concentration meteorological classification tree

表1　决策树模型识别气象分类结果

Tab. 1　Classification for regional meteorological

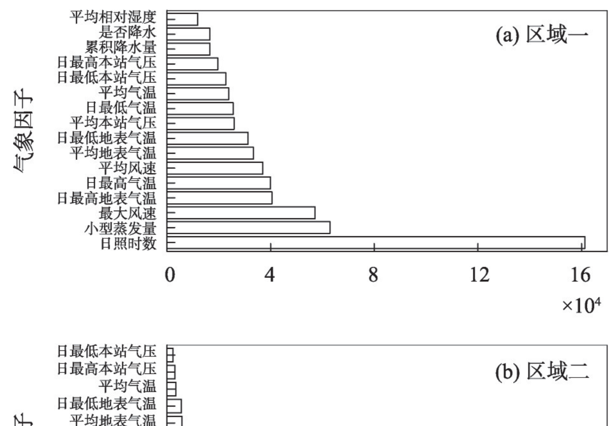
类别	气象参数	个数/个	占比/%	$PM_{2.5}$平均浓度值/(μg·m^{-3})
	区域一			
Ⅰ-2	日照时数≥7.88 h	122	26.93	51.75
Ⅰ-5	日照时数<7.88 h；最大风速≥3.72 m·s^{-1}；日最低地表气温≥-0.10 ℃	65	14.35	58.73
Ⅰ-7	日照时数<7.88 h；最大风速≥3.72 m·s^{-1}；日最低地表气温<-0.10 ℃；平均本站气压≥919.53 hPa	33	7.28	60.74
Ⅰ-8	日照时数<7.88 h；最大风速≥3.72 m·s^{-1}；日最低地表气温<-0.10 ℃；平均本站气压<919.53 hPa	108	23.84	92.75
Ⅰ-10	1.57 h≤日照时数<7.88 h；最大风速<3.72 m·s^{-1}	75	16.56	87.85
Ⅰ-11	日照时数<1.57 h；最大风速<3.72 m·s^{-1}	50	11.04	128.20
	区域二			
Ⅱ-3	小型蒸发量≥0.96 mm；平均相对湿度<45.38%	95	20.97	66.59
Ⅱ-5	小型蒸发量≥0.96 mm；平均相对湿度≥45.38%；日照时数≥8.55 h	52	11.48	77.44
Ⅱ-7	小型蒸发量≥0.96 mm；平均相对湿度≥45.38%；日照时数<8.55 h；累积降水量≥0.11 mm	39	8.61	77.82
Ⅱ-9	小型蒸发量≥0.96 mm；平均相对湿度≥45.38%；日照时数<8.55 h；累积降水量<0.11 mm；平均风速<2.43 m·s^{-1}	98	21.63	112.72
Ⅱ-10	小型蒸发量≥0.96 mm；平均相对湿度≥45.38%；日照时数<8.55 h；平均风速≥2.43 m·s^{-1}	90	19.87	142.18
Ⅱ-11	小型蒸发量<0.96 mm	79	17.44	171.26

情况下主要是由湿度、降水的变化而导致的波动。通过对所有参与模型的气象因子重要程度进行分析(图6)，可以看出，区域一二所选择的模型有效涵盖了相关重要因子，所识别的关键气象因子在胡琳等[28]的研究中也得到了有效支持。

为了有效验证所构建分类模型的可信度。以区域一为例，对基于$PM_{2.5}$分类后的气象类别再次采用决策树预测模型进行验证分类结果的准确性。通过选取70%的训练集和30%的测试集对分类结果进行验证，测试集预测结果的准确率达到了81.2%，有效证明了该方法的可行性。

2.4　最不利气象条件识别

将上述的气象分类与$PM_{2.5}$的浓度相结合绘制时间序列图7。由图7可以看出，对于区域一而言，$PM_{2.5}$的浓度呈由气象条件驱动下的周期性变化。在气象Ⅰ-2、Ⅰ-5、Ⅰ-7条件下，$PM_{2.5}$基本处于低浓度状态；在1月中旬至3月期间$PM_{2.5}$的浓度周期性波动主要是Ⅰ-8类气象条件所导致，在Ⅰ-8类气象条件下，区域一$PM_{2.5}$总体呈下降趋势，在清除$PM_{2.5}$污染。在12月至2月中旬主要是Ⅰ-10和Ⅰ-11类



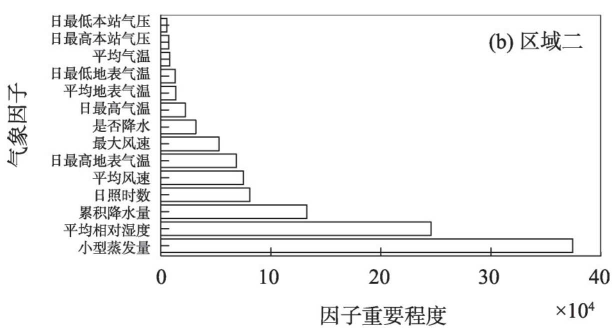


图6　气象因子重要度分析

Fig. 6　Importance analysis of meteorological factors

气象条件驱动的$PM_{2.5}$浓度变化,这两类气象条件所驱动的$PM_{2.5}$浓度峰值会高于在Ⅰ-8类气象条件下的峰值。对于区域二而言,在Ⅱ-3、Ⅱ-5、Ⅱ-7气象条件下,$PM_{2.5}$基本处于低浓度状态。在Ⅱ-9类气象条件下,区域二$PM_{2.5}$总体呈下降趋势,在清除$PM_{2.5}$污染。在11月至12月中旬和2月中旬至3月期间主要是由Ⅱ-10类气象条件所驱动,12月中旬至2月中旬主要是由Ⅱ-11类气象条件所驱动导致$PM_{2.5}$浓度周期性变化。由Ⅱ-11类气象条件所驱动的$PM_{2.5}$浓度周期性变化峰值会高于由Ⅱ-10类气象条件所驱动的峰值。

综上所述,区域一$PM_{2.5}$高浓度主要发生在Ⅰ-10和Ⅰ-11气象条件下;区域二的$PM_{2.5}$高浓度主要发生在Ⅱ-10和Ⅱ-11类气象条件。将此类气象条件标记为不利气象条件,在这些气象条件下,$PM_{2.5}$环境容量会达到区域最小值。故需要对此单独进行研究。

2.5　$PM_{2.5}$浓度变化

由于区域一中Ⅰ-10和Ⅰ-11两类气象条件比较相近,将区域一中Ⅰ-10和Ⅰ-11两类气象条件合并。选取区域一和区域二由Ⅰ-10、Ⅰ-11和Ⅱ-10、Ⅱ-11开始的气象条件,连续3 d及以上时间$PM_{2.5}$浓度持续上涨;其中为了保证选取时间段的连续性,有个别时间段有1 d的其他气象类型或浓度下降的情况。最终区域一有13个时间段满足条件,初始平均浓度为90.02 $\mu g\cdot m^{-3}$,每段平均持续4.92 d;区域二Ⅱ-10有11个时间段满足条件,初始平均浓度为108.26 $\mu g\cdot m^{-3}$,每段平均持续4.22 d。区域二Ⅱ-11有7个时间段满足条件,初始平均浓度为182.71 $\mu g\cdot m^{-3}$,每段平均持续5.14 d。

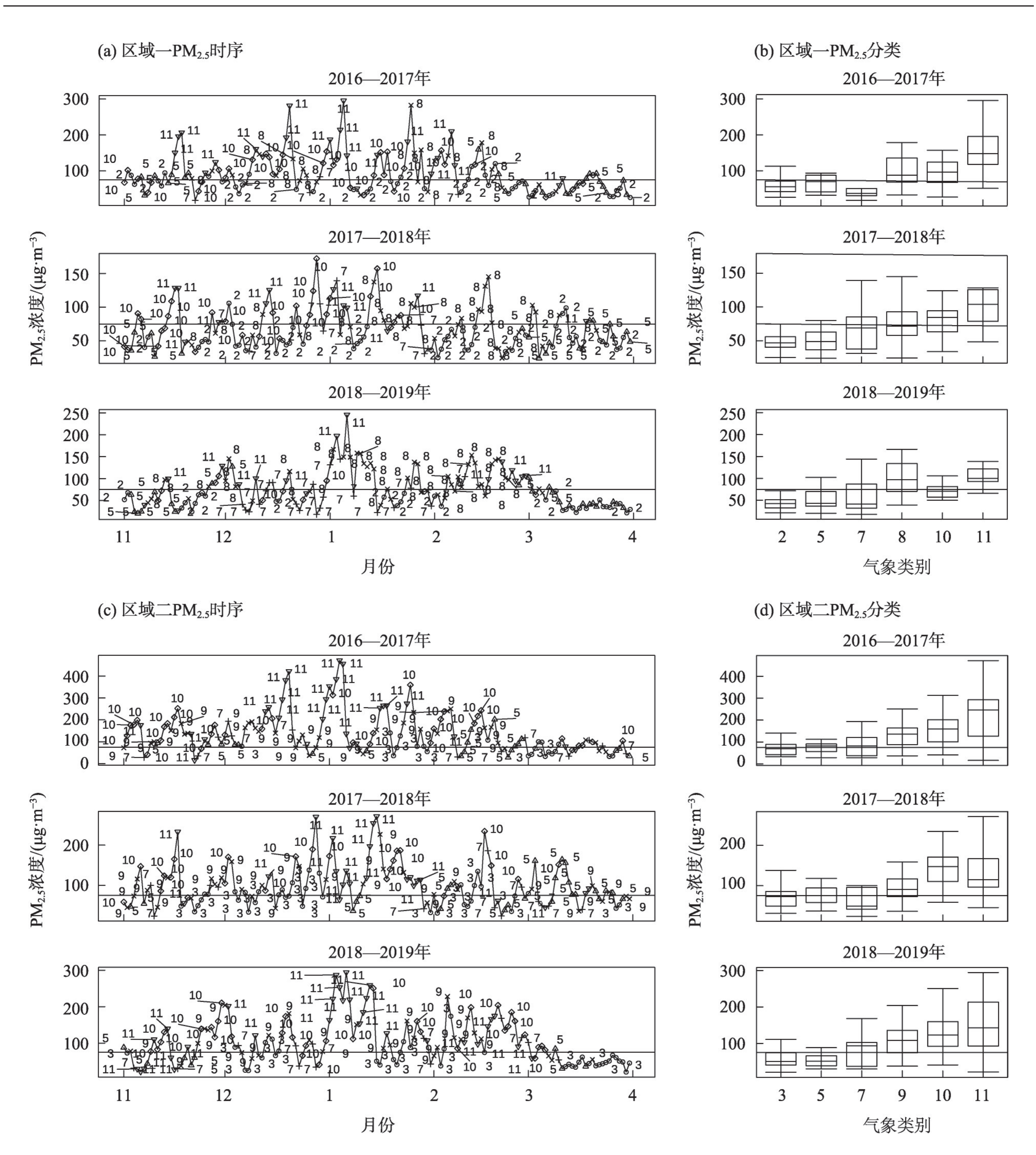
对区域一和区域二所选取的每一段进行回归构建最不利天气下$PM_{2.5}$浓度的增长模型(图8)。图中标注了$PM_{2.5}$浓度累积最快和最缓的回归直线方程。根据以上信息,得出区域一在Ⅰ-10、Ⅰ-11最不利气象条件下,$PM_{2.5}$浓度的平均变化方程为:$y=21.693x+90.02$(图8a);区域二在Ⅱ-10最不利气象条件下,$PM_{2.5}$浓度的平均变化方程为:$y=18.746x+108.26$(图8b)。区域二在Ⅱ-11最不利气象条件下,$PM_{2.5}$浓度的平均变化方程为:$y=27.658x+182.71$(图8c)。

由表1可知,区域一在Ⅰ-2、Ⅰ-5、Ⅰ-7气象类型,区域二Ⅱ-3、Ⅱ-5、Ⅱ-7气象类型下,$PM_{2.5}$平均浓度基本满足环境空气质量标准(GB3095-2012)中规定的$PM_{2.5}$日均值浓度75 $\mu g\cdot m^{-3}$二级标准。而区域一Ⅰ-10、Ⅰ-11类和区域二Ⅱ-9、Ⅱ-10类最不利气象条件下,区域一、二$PM_{2.5}$浓度远超二级标准。

回归结果显示,在区域一Ⅰ-10、Ⅰ-11类和区域二Ⅱ-9、Ⅱ-10类最不利气象条件下,$PM_{2.5}$最高浓度分别超GB3095-2012中24 h平均$PM_{2.5}$(标准数值75 $\mu g\cdot m^{-3}$)二级标准的2.62倍,2.50倍和4.33倍。可以看出,相关区域出现所识别的最不利气象条件以及初始浓度达到一定程度时,相关主管部门应按不同的控制程度对企业污染排放设定时间段。由于$PM_{2.5}$浓度与污染排放之间呈现的是非线性关系[29-30],且不同的工业源、移动源、农业面源、居民源等排放在不同的气象条件和不同区域下对$PM_{2.5}$浓度的贡献不同[31-33]。

3　结论与建议

(1) 关中地区根据冬防期日尺度$PM_{2.5}$浓度均值结合地形数据,对空间聚类结果进行相关系数计算聚类后可划分为2个污染区域:关中平原等低海拔地形区域和海拔相对较高的山脉区域。关中地区冬防期$PM_{2.5}$平均浓度随着海拔的升高而降低,当海拔升至100 m左右后,$PM_{2.5}$日均浓度值逐渐趋于



注：图中的横线对应浓度为环境空气质量标准(GB3095-2012)中规定的$PM_{2.5}$日均值二级标准75 μg·m^{-3}。

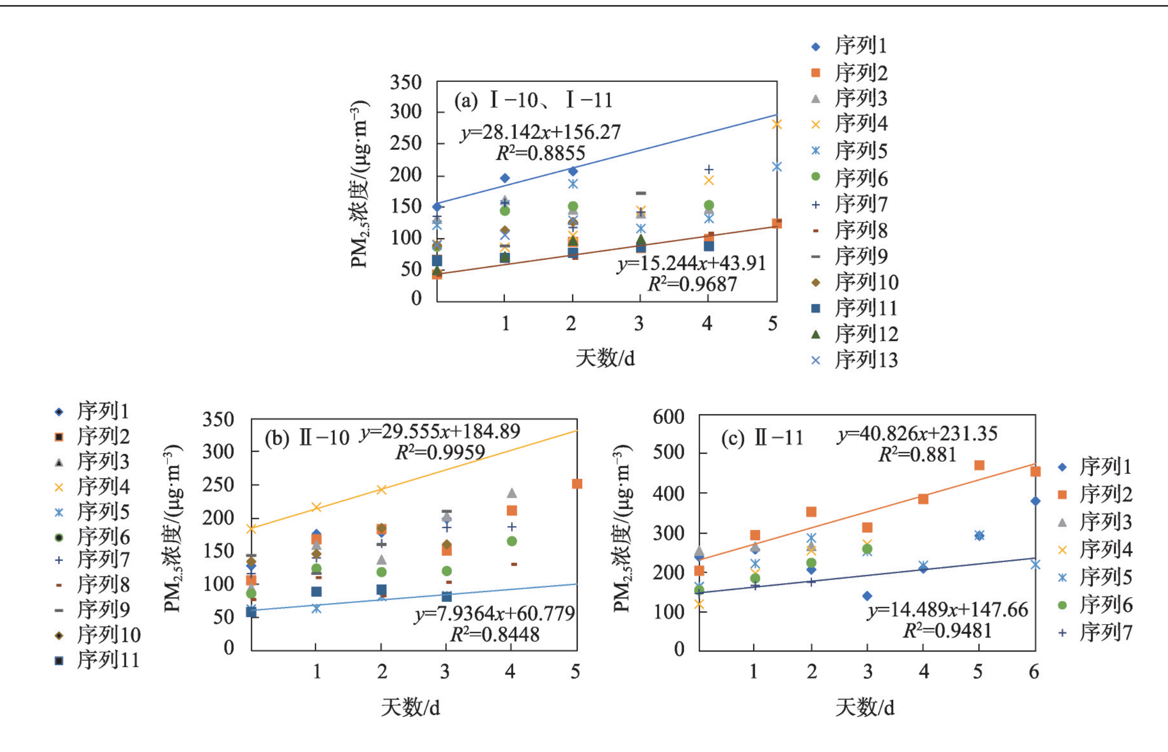
图a、图c中标注的数字为$PM_{2.5}$所对应的气象类别。

图7　$PM_{2.5}$浓度与气象类型时序

Fig. 7　$PM_{2.5}$ concentration and meteorological type sequence diagram

稳定。

（2）对于分区后的地区分别采用决策树对其气象类型基于$PM_{2.5}$浓度进行分类，其中高海拔地区在模型构建的时候选取了日照时数、最大风速、日最低地表气温、平均本站气压4项指标将气象分为6类。低海拔地区在模型构建的时候选取小型蒸发量、平均相对湿度、日照时数、累积降水量、平均风速5项指标将气象分为6类。高海拔区域在Ⅰ-10和Ⅰ-11两类气象条件的$PM_{2.5}$浓度较高；低海拔区域在Ⅱ-10和Ⅱ-11两类气象条件的$PM_{2.5}$浓度较高；高海拔和低海拔区域分别在Ⅰ-8类、Ⅰ-9类气象条件下的$PM_{2.5}$浓度呈下降趋势。



注:图a为区域一在Ⅰ-10、Ⅰ-11最不利气象条件下$PM_{2.5}$浓度平均变化方程;图b为区域二在Ⅱ-10最不利气象条件下$PM_{2.5}$浓度平均变化方程;图c为区域二在Ⅱ-11最不利气象条件下$PM_{2.5}$浓度平均变化方程。

图8　区域一、二最不利气象条件下$PM_{2.5}$回归模型

Fig. 8　$PM_{2.5}$ regression model under the most unfavorable meteorological conditions in region Ⅰ and Ⅱ

(3)回归分析显示,高海拔区域在Ⅰ-10、Ⅰ-11类和低海拔区域二Ⅱ-10、Ⅱ-11类最不利气象条件下,$PM_{2.5}$浓度平均会持续上升4.76 d,到达最高浓度。最高浓度分别超GB3095-2012中24 h平均$PM_{2.5}$二级标准的2.62倍,2.50倍和4.33倍。

(4)建议在某一区域发生所识别的不利气象条件时,应对该区域内城市同时发布重污染天气预警要求。为有效减少社会经济损失和对居民身体健康的伤害等,发布预警时应包含可能污染持续时长及可能的严重程度等信息,从而促使居民提前安排出行,污染企业提前做好生产任务安排。

参考文献(References):

[1] Guo Y M, Jia Y P, P X C, et al. The association between fine particulate air pollution and hospital emergency room visits for cardiovascular diseases in Beijing, China[J]. Science of the Total Environment, 2009, 407(17): 4826-4830.

[2] Guo P, Feng W R, Zheng M R, et al. Short-term associations of ambient air pollution and cause-specific emergency department visits in Guangzhou, China[J]. Science of the Total Environment, 2018, 613-614: 306-313.

[3] 张增凯, 路雨婷, 赵鹏宇, 等. 2019年京津冀重污染天气应急响应经济损失评估[J]. 中国环境科学, 2021, 41(7): 3399-3408. [Zhang Zengkai, Lu Yuting, Zhao Pengyu, et al. The assessment of economic cost induced by emergency responses for heavy pollution weather within Jing-Jin-Ji area in 2019[J]. China Environmental Science, 2021, 41(7): 3399-3408.]

[4] Nie T, Nie L, Zhou Z, et al. Exploring the heavy air pollution in Beijing in the fourth quarter of 2015: Assessment of environmental benefits for red alerts[J]. Frontiers of Earth Science, 2017, 12(2): 361-372.

[5] Zhao H P, Wang F P, Niu C C, et al. Red warning for air pollution in China: Exploring residents' perceptions of the first two red warnings in Beijing[J]. Environmental Research, 2018, 161: 540-545.

[6] Tribby C P, Miller H J, Song Y, et al. Do air quality alerts reduce traffic? An analysis of traffic data from the Salt Lake City metropolitan area, Utah, USA[J]. Transport Policy, 2013, 30: 173-185.

[7] Saberian S, Heyes A, Rivers N. Alerts work! Air quality warnings and cycling[J]. Resource and Energy Economics, 2017, 49: 165-185.

[8] Zivin J G, Neidell M. Days of haze: Environmental information disclosure and intertemporal avoidance behavior[J]. Journal of Environmental Economics and Management, 2009, 58(2): 119-128.

[9] Neidell M J. Air pollution, health, and socio-economic status: the effect of outdoor air quality on childhood asthma[J]. Journal of Health Economics, 2004, 23(6): 1209-1236.

[10] Troncoso R, Grange L D, Cifuentes L A. Effects of environmental

alerts and pre-emergencies on pollutant concentrations in Santiago, Chile[J]. Atmospheric Environment, 2012, 61: 550–557.

[11] 王力, 姜晓群. 重污染天气应急预警下的固定源"一厂一策"实施机制探讨[J]. 环境保护, 2019, 47(7): 53–56. [Wang Li, Jiang Xiaoqun. Research on the implementation mechanism of One Factory & One Policy at stationary source in heavy air pollution days [J]. Environmental Protection, 2019, 47(7): 53–56.]

[12] 贺璇. 技术赋能与空气重污染应急管理智慧化演进路径[J]. 华中科技大学学报(社会科学版), 2020, 34(3): 131–140. [He Xuan. Starting Characteristics of air pollution emergency management and intelligent evolution by technology empowerment[J]. Journal of Huazhong University of Science and Technology (Social Science Edition), 2020, 34(3): 131–140.]

[13] 李名升, 任晓霞, 于洋, 等. 中国大陆城市$PM_{2.5}$污染时空分布规律[J]. 中国环境科学, 2016, 36(3): 641–650. [Li Mingsheng, Ren Xiaoxia, Yu Yang, et.al. Spatiotemporal pattern of ground-level fine particulate matter ($PM_{2.5}$) pollution in mainland China[J]. China Environmental Science, 2016, 36(3): 641–650.]

[14] 徐大海, 王郁, 朱蓉. 大气环境容量系数A值频率曲线拟合及其应用[J]. 中国环境科学, 2016, 36(10): 2913–2922. [Xu Dahai, Wang Yu, Zhu Rong. The atmospheric environmental capacity coefficient cumulative frequency curve fitting and its application[J]. China Environmental Science, 2016, 36(10): 2913–2922.]

[15] 朱蓉, 张存杰, 梅梅. 大气自净能力指数的气候特征与应用研究[J]. 中国环境科学, 2018, 38(10): 3601–3610. [Zhu Rong, Zhang Cunjie, Mei Mei. The climate characteristics of atmospheric self-cleaning ability index and its application in China[J]. China Environmental Science, 2018, 38(10): 3601–3610.]

[16] Wang J Z, Wang Y Q, Liu H, et al. Diagnostic identification of the impact of meteorological conditions on $PM_{2.5}$ concentrations in Beijing[J]. Atmospheric Environment, 2013, 81:158–165.

[17] Yang J H, Ji Z M, Kang S C, et al. Spatiotemporal variations of air pollutants in western China and their relationship to meteorological factors and emission sources[J]. Environmental Pollution, 2019, 254(Pt A): 112952.

[18] 高爽, 李时蓓, 马岩, 等. 基于气象分型的大气环境允许排放量测算方法[J]. 中国环境科学, 2021, 41(2): 588–595. [Gao Shuang, Li Shibei, Ma Yan, et al. Research on atmospheric environmental capacity accounting method based on classification of meteorological conditions[J]. China Environmental Science, 2021, 41 (2): 588–595.]

[19] 许艳玲, 薛文博, 雷宇. 气象和排放变化对$PM_{2.5}$污染的定量影响[J]. 中国环境科学, 2019, 39(11): 4546–4551. [Xu Yanlin, Xue Wenbo, Lei Yu. Impact of meteorological conditions and emission change on $PM_{2.5}$ pollution in China[J]. China Environmental Science, 2019, 39(11): 4546–4551.]

[20] 兰紫娟, 卢超, 李磊, 等. 基于污染气象条件分级的深圳$PM_{2.5}$污染回顾评价[J]. 环境科学学报, 2020, 40(12): 4410–4418. [Lan Zijuan, Lu Chao, Li Lei, et al. Review of $PM_{2.5}$ pollution in Shenzhen based on classification of pollution meteorological condition [J]. Acta Scientiae Circumstantiae, 2020, 40(12): 4410–4418.]

[21] Xue Y L, Xue W B, Lei Y, et al. Impact of meteorological conditions on $PM_{2.5}$ pollution in China during winter[J]. Atmosphere, 2018, 9(11): 429.

[22] 郝志新, 耿秀, 刘可邦, 等. 关中平原过去1000年干湿变化特征[J]. 科学通报, 2017, 62(21): 2399–2406. [Hao Zhixin, Geng Xiu, Liu Kebang, et al. Dryness and wetness variations for the past 1000 years in Guanzhong Plain[J]. Chinese Science Bulletin, 2017, 62(21): 2399–2406.]

[23] 陈建文, 胡琳, 王娟敏, 等. 陕西省大气稳定度分布特征研究[J]. 水土保持研究, 2013, 20(3): 299–304. [Chen Jianwen, Hu Lin, Wang Juanmin, et al. Distribution characteristics of atmospheric stability in Shaanxi Province[J]. Research of Soil and Water Conservation, 2013, 20(3): 299–304.]

[24] Brownlee J. Machine learning mastery[J]. URL: http://machinelearningmastery.com/discover-feature-engineering-howtoengineer-features-and-how-to-getgood-at-it, 2014.

[25] 贾册, 王亚琼, 杨啸. 关中平原$PM_{2.5}$浓度周期性变化规律及成因分析[J]. 干旱区研究, 2020, 37(6): 1528–1536. [Jia Ce, Wang Yaqiong, Yang Xiao. Analysis on the periodic variation of $PM_{2.5}$ concentration and its causes in Guanzhong Plain[J]. Arid Zone Research, 2020, 37(6): 1528–1536.]

[26] 张波, 宋国君, 周芳. 基于$PM_{2.5}$监测点空间聚类的关中五市空气污染区域识别[J]. 环境科学学报, 2021, 41(3): 1–9. [Zhang Bo, Song Guojun, Zhou Fang. Regional delimitation of $PM_{2.5}$ pollution: A case study of five cities in Guanzhong Plain[J]. Acta Scientiae Circumstantiae, 2021, 41(3): 1–9.]

[27] 宋国君, 国潇丹, 杨啸, 等. 沈阳市$PM_{2.5}$浓度ARIMA-SVM组合预测研究[J]. 中国环境科学, 2018, 38(11): 4031–4039. [Song Guojun, Guo Xiaodan, Yang Xiao, et.al. ARIMA-SVM combination prediction of $PM_{2.5}$ concentration in Shenyang[J]. China Environmental Science, 2018, 38(11): 4031–4039.]

[28] 胡琳, 程路, 林扬, 等. 关中区域大气重污染年份气象条件分析[J]. 干旱区研究, 2020, 37(6): 1496–1503. [Hu Lin, Cheng Lu, Lin Yang, et al. A conceptual model of regional pollution and meteorological conditions in Guanzhong region[J]. Arid Zone Research, 2020, 37(6): 1496–1503.]

[29] Zhao B, Wang S X, Ding D, et al. Nonlinear relationships between air pollutant emissions and $PM_{2.5}$-related health impacts in the Beijing-Tianjin-Hebei region[J]. Science of The Total Environment, 2019, 661: 375–385.

[30] Zhao B, Wang S X, Xing J, et al. Assessing the nonlinear response of fine particles to precursor emissions: Development and application of an extended response surface modeling technique v1.0[J]. Geoscientific Model Development, 2015, 8(1): 115–128.

[31] 王静晞, 曹国良, 韩蕾, 等. 关中地区人为源大气污染物排放清单研究[J]. 安全与环境学报, 2015, 15(5): 282–287. [Wang Jingxi, Cao Guoliang, Han Lei, et al. Probe to the anthropogenic

air pollutant emission inventory in Guanzhong Region, Shaanxi[J]. Journal of Safety and Environment, 2015, 15(5): 282−287.]

[32] 郭庆元, 杨晓春, 吴其重, 等. 基于数值模拟对西安地区冬季一次重污染过程 $PM_{2.5}$ 区域来源解析[J]. 环境科学学报, 2020, 40(9): 3103−3111. [Guo Qingyuan, Yang Xiaochun, Wu Qizhong, et al. An analysis of $PM_{2.5}$ regional sources from heavy pollution in the winter of Xi'an based on numerical simulation[J]. Acta Scientiae Circumstantiae, 2020, 40(9): 3103−3111.]

[33] 李霞, 吴佳睿, 刘浪, 等. 模拟研究咸阳市2016年秋冬季雾霾期间 $PM_{2.5}$ 来源[J]. 地球环境学报, 2019, 10(4): 347−363. [Li Xia, Wu Jiarui, Liu Lang, et al. Simulating the sources of $PM_{2.5}$ during heavy haze pollution episodes in the autumn and winter of 2016 in Xianyang City, China[J]. Journal of Earth Environment, 2019, 10(4): 347−363.]

Optimal time period for $PM_{2.5}$ control based on decision tree model: A case study of Guanzhong, China

JIA Ce[1], CHEN Zhen[2,3], HAN Mei[4]

(1. School of Environment & Nature Resources, Renmin University of China, Beijing 100872, China; 2. Institutes of Science and Development, Chinese Academy of Sciences, Beijing 100190, China; 3. School of Public Policy and Management, University of Chinese Academy of Sciences, Beijing 100049, China; 4. Shaanxi Provincial Investigation and Ecological Assessment Center, Xi'an 710054, Shaanxi, China)

Abstract: A simple spatiotemporal identification method for $PM_{2.5}$ pollution control based on a machine learning model is proposed. A spatiotemporal clustering algorithm was used to cluster $PM_{2.5}$ in the winter prevention period, and different $PM_{2.5}$ polluted areas were identified. Furthermore, a decision tree model was constructed using the meteorological data of different areas to identify the most unfavorable influences on $PM_{2.5}$ concentration in different areas. The changes in $PM_{2.5}$ under the most unfavorable meteorological conditions were analyzed, and the optimal time period for $PM_{2.5}$ pollution control in the different study areas was determined. Using the Guanzhong area as an example, the correlation analysis of the spatial clustering of mean daily $PM_{2.5}$ in winter showed that the Guanzhong area is mainly divided into a low-altitude plain area (the Guanzhong Plain) and a relatively high mountain range. The classification tree model was then constructed, using the meteorological data of the plain and mountain areas. The analysis showed that the mountain elevations have a higher $PM_{2.5}$ concentrations under Ⅰ−10 (1.57 h ≤ sunshine hours < 7.88 h; maximum wind speed < 3.72 $m \cdot s^{-1}$) and Ⅰ−11 (sunshine hours < 1.57 h; maximum wind speed < 3.72 $m \cdot s^{-1}$) meteorological conditions; low-altitude areas have a higher $PM_{2.5}$ concentrations under Ⅱ−10 (small-scale evaporation ≥ 0.96 mm; average relative humidity ≥ 45.38%; sunshine hours < 8.55 h; average wind speed ≥ 2.43 $m \cdot s^{-1}$) and Ⅱ−11 (small evaporation < 0.96 mm) meteorological conditions. Lastly, the regression analysis results showed that in the mountain elevations area in the Ⅰ−10, Ⅰ−11 category and the plain altitude area under the most unfavorable meteorological conditions (Ⅱ−10 and Ⅱ−11 category), the $PM_{2.5}$ concentration will continue to increase for 4.76 days on average, before reaching a maximum concentration.

Keywords: decision tree model; $PM_{2.5}$; meteorological classification; heavy pollution